BÀI TẬP 1
THỐNG KÊ MÁY TÍNH VÀ ỨNG DỤNG
KÌ 2 2021-2022, ĐÀO TẠO TỪ XA

Lưu ý: Nộp file Python Notebook. Ghi rõ thông tin sinh viên ở đầu file.

Câu 1. (2.5 điểm) Bảng dưới đây cho thấy độ tuổi của những người nghiện ma túy tại một trung tâm cai nghiện vào các năm 1971, 1972.

	1971	**1972**
Dưới 20	118	96
Từ 20 – dưới 25	772	727
Từ 25 – dưới 30	288	376
Từ 30 – dưới 35	112	117
Từ 35 – dưới 50	112	118
Trên 50	177	165

a) Vẽ biểu đồ quạt (pie chart) cho năm 1971.
b) Vẽ biểu đồ thanh (bar chart) cho năm 1972.
c) Vẽ biểu đồ so sánh 2 năm.
d) Nhận xét.

Câu 2. (2.5 điểm) Bảng sau là kết quả khảo sát lượng thuốc hút trung bình mỗi ngày của 68 người có hút thuốc.

Số điếu thuốc hút trung bình 1 ngày	[0, 8)	[8, 12)	[12, 16)	[16, 24)	[24, 28)	[28, 34)	[34, 50]
Số lượng	4	6	12	28	8	6	4

a) Lập bảng cho thấy tần số (frequency), tần suất (relative frequency), mật độ (density), tần số tích lũy (cumulative frequency), tần suất tích lũy (cumulative relative frequency).
b) Vẽ biểu đồ histogram.
c) Vẽ biểu đồ tần suất tích lũy.
d) Nhận xét.

Câu 3. (2.5 điểm) Sau đây là số liệu lượng hành khách trên 50 chuyến tàu được khảo sát

165 141 163 153 130 158 119 187 185 209

177 147 166 154 159 178 187 139 180 143

160 185 153 168 189 173 127 179 163 182

171 146 174 149 126 156 155 174 154 150

210 162 138 117 198 164 125 142 182 218

a) Tính trung bình (mean), trung vị (median), yếu vị (mode), phân vị (percentile) 10%, 90%.

b) Tính phạm vi (range), khoảng tứ phân vị (IQR - interquartile range), phương sai (variance), độ lệch chuẩn (standard deviation).
c) Nhận xét.

Câu 4. (2.5 điểm) Tập tin `WHO-COVID-19-global-data.csv` chứa dữ liệu số lượng ca nhiễm và chết do COVID-19 của các nước từ đầu năm 2020 đến cuối năm 2021 (nguồn WHO: https://covid19.who.int/WHO-COVID-19-global-data.csv).

a) "Trích và mô tả" dữ liệu của Việt Nam (mã quốc gia là VN).
b) "So sánh" dữ liệu của Việt Nam với Thái Land (mã quốc gia là TH).

--- HẾT ---